An Axiomatic Foundation for Social Learning in Networks

Ali Jadbabaie

Institute for Data, Systems, and Society, MIT
Alfred Fitler Moore Professor
Electrical and Systems Engineering, Computer & Information Sciences
Operations, Information and Decisions
University of Pennsylvania

with Pooya Molavi (MIT Economics), Alireza Tahbaz Salehi(Columbia GSB),
M. Amin Rahimian (Penn)

Social Learning and Distributed Inference

- How do individuals aggregate opinions/actions?
- What is the influence of private observations and their social interactions?
- No central mechanisms for aggregation
- Examples: purchasing products, suitability of political candidates, opinion on climate change,purchase of new product,...
- Long history in economics and engineering

How does information and network structures influence learning?

Social Learning

- $\{1, \ldots, n\}$: finite set of agents
- Agents want to learn an underlying state $\theta \in \Theta$.
- $t \in \mathbb{N}$: discrete time
- The state is drawn at $t = 0$ according to agents' common prior.
- $\omega_{it} \in S$: private observations of agent i at time t
- Conditional on θ being realized, $\omega_{it} \sim \ell_i^\theta \in \Delta S$.
- $\ell_i = \{\ell_i^\theta\}_{\theta \in \Theta}$: agent i's signal structure

Social Learning

- $\{1, \ldots, n\}$: finite set of agents
- Agents want to learn an underlying state $\theta \in \Theta$.
- $t \in \mathbb{N}$: discrete time
- The state is drawn at $t = 0$ according to agents' common prior.
- $\omega_{it} \in S$: private observations of agent i at time t
- Conditional on θ being realized, $\omega_{it} \sim \ell_i^\theta \in \Delta S$.
- $\ell_i = \{\ell_i^\theta\}_{\theta \in \Theta}$: agent i's signal structure

Assumption (identifiability)

For all $\theta, \hat{\theta} \in \Theta$, there exists i such that $\ell_i^\theta \neq \ell_i^{\hat{\theta}}$.

Standard Bayesian Learning

- Bayesian update:

$$\mu_{it+1}(\theta) = \mathrm{BU}(\mu_{it}(\theta), \omega_{it+1}) := \frac{\mu_{it}(\theta)\ell_i^{\theta}(\omega_{it+1})}{\sum_{\tilde{\theta}\in\Theta} \mu_{it}(\tilde{\theta})\ell_i^{\tilde{\theta}}(\omega_{it+1})}$$

- Rate = $\min_{\check{\theta}} h_i(\theta, \check{\theta}) := D_{KL}\left(\ell_i^{\theta}(\cdot)\|\ell_i^{\theta^*}(\cdot)\right)$.

- A stochastic gradient descent algorithm applied to max-likelihood

$$\max_{\mu\in\Delta\Theta}\left\{\mu^T \mathbb{E}^*_{\omega_t}[\log \ell_i^{\theta}(\omega_{t+1})] + D_{KL}(\mu\|\mu_0)\right\}.$$

The Bayesian Benchmark: Multiagent setting

- Let $\mathcal{X} = \overbrace{\Theta}^{\text{state}} \times \overbrace{\Omega}^{\text{signals}} \times \overbrace{\Gamma}^{\text{network}}$ be the measurable space that captures *all* uncertainty.
- Assume agents have a *common* prior over the $\mathcal{X}$.

The Bayesian Benchmark: Multiagent setting

- Let $\mathcal{X} = \underbrace{\Theta}^{\text{state}} \times \overbrace{\Omega}^{\text{signals}} \times \overbrace{\Gamma}^{\text{network}}$ be the measurable space that captures *all* uncertainty.
- Assume agents have a *common* prior over the $\mathcal{X}$.

Theorem

(J., Molavi, Sandroni, Tahbaz-Salehi, 2012), Assume

1. *agents' common prior has full support over $\mathcal{X}$;*
2. *the realized network is strongly connected;*
3. *the realized state is identifiable.*

Then all agents learn the true state asymptotically almost surely; i.e.,

$$\mu_{it} \longrightarrow \mathbf{1}_{\theta^*} \qquad \forall i \in \mathcal{N}.$$

also, Mossel, Sly and Tamuz (2015) for forward-looking agents

Non-Bayesian Models

Full Bayesian learning is intractable

What to do?

- DeGroot (1974)
- Ellison and Fudenberg (1993, 1995)
- Bala and Goyal (1998, 2001)
- DeMarzo, Vayanos, and Zwiebel (2003)
- Acemoglu, Ozdaglar, and Parandeh-Gheibi (2010)
- Golub and Jackson (2010)
- J, Molavi, Sandroni, Tahbaz-Salehi (2012)

Non-Bayesian Learning

Source of difficulty of Bayesian: History dependence

- Empirical evidence for DeGroot-like models (Chandrasekhar et al. 2015).
- Bayesian update uses entire history.
- Sequential Bayesian updates can exhibit anti-imitative behavior (Eyster & Rabin 2010, 2014)
- DeGroot models neglect history

Our solution (2012-2014):

DeGroot-like models with

- Continuous flow of new information with heterogenous observations
- Asymptotic agreement with the Bayesian update
- Are these non-Bayesian updates adhoc? Is there a foundation?
- How do these updates depart from Bayesian?

Bayesian Learning Without Recall (BWR)

- What if agents don't recall how others form their beliefs?
- Can we circumvent history dependence?
- Difficulty starts after time step 1. What if we copy that map for all times?
- Will agents learn?

Replicate the Bayesian map of time 0 to 1, for all future time steps

BWR update is log-linear

- Take $\Theta = \{\theta_1, \ldots, \theta_m\}$, $\mathcal{A}_i = \Delta\Theta$ and $u_i(\bar{a}, \theta_j) = -\|\bar{a} - \bar{e}_j\|_2^2$.
- $\mathbf{a}_{i,t} = \arg\max_{a \in \mathcal{A}_i} \mathbb{E}_{i,t}\{u_i(a, \boldsymbol{\theta})\} \equiv \boldsymbol{\mu}_{i,t}(\cdot)$.

$\downarrow$

$$\boldsymbol{\mu}_{it+1}(\check{\theta}) = \frac{\nu_i(\check{\theta})\ell_i^{\check{\theta}}(\omega_{it+1})\left(\prod_{j\in\mathcal{N}(i)} \frac{\boldsymbol{\mu}_{jt}(\check{\theta})}{\nu_j(\check{\theta})}\right)}{\sum_{\tilde{\theta}\in\Theta} \nu_i(\tilde{\theta})\ell_i^{\tilde{\theta}}(\omega_{it+1})\left(\prod_{j\in\mathcal{N}(i)} \frac{\boldsymbol{\mu}_{jt}(\tilde{\theta})}{\nu_j(\tilde{\theta})}\right)}$$

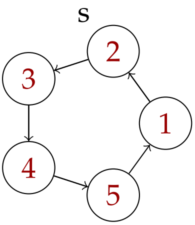
Theorem

Rahimian, J (2014) *No agent can learn the truth almost surely unless $\rho(A) = 1$.*

- Directed cycles, rooted trees with degree 1.
- if $\rho(A) > 1$, misinformation gets amplified.
- Neighbors beliefs *not* sufficient statistics!

Learning with $\rho(A) = 1$

$$\boldsymbol{\mu}_{i,t}(\check{\theta}) = \frac{\ell_i(\mathbf{s}_{i,t}|\check{\theta})\boldsymbol{\mu}_{j,t-1}(\check{\theta})}{\sum_{\tilde{\theta}\in\Theta} \ell_i(\mathbf{s}_{i,t}|\tilde{\theta})\boldsymbol{\mu}_{j,t-1}(\tilde{\theta})}, \mathcal{N}(i) = \{j\}.$$



Rate of Learning: $\frac{1}{n}\sum_{i=1}^{n} D_{KL}\left(\ell_i(\cdot|\theta)\|\ell_i(\cdot|\check{\theta})\right)$.

For a general strongly connected topology:

- Choose neighbor $j \in \mathcal{N}(i)$ independently at random w.p. $[P]_{i,j}$.
- $(\pi_1, \ldots, \pi_n)$ is the stationary distribution for P, irreducible.
- Asymptotic Rate: $\sum_{i=1}^{n} \pi_i D_{KL}\left(\ell_i(\cdot|\theta)\|\ell_i(\cdot|\check{\theta})\right)$.

Deriving Linear Majority Rules

- Take $\mathcal{A}_i = \Theta = \{\pm 1\}$, $u_i(a,\theta) = 2\mathbb{1}_{\theta=a} - 1$.
- $\mathbf{a}_{i,t} = \arg\max_{a \in \mathcal{A}_i} \mathbb{E}_{i,t}\{u_i(a,\theta)\} = \text{sign}\left(\log\left(\boldsymbol{\mu}_{i,t}(+1)/\boldsymbol{\mu}_{i,t}(-1)\right)\right).$

$$\downarrow$$

$$\mathbf{a}_{i,t} = \text{sign}\left(\sum_{j \in \mathcal{N}(i)} w_j \mathbf{a}_{j,t-1} + \eta_i + \lambda(\mathbf{s}_{i,t})\right)$$

- Action profiles evolve as a Markov chain on the Boolean cube.
- Consensus is an equilibrium $\leftrightarrow \max_{s_i \in \mathcal{S}_i} |\lambda(s_i) + \eta_i| < \sum_{j \in \mathcal{N}(i)} w_j$.
- With a positive probability the agents (mis-)learn.

(log)-linear Learning Rules

DeGroot+Bayesian: (log)-linear aggregation

- Agents aggregate the observed beliefs of their neighbors (log)-linearly and process their private observations in a Bayesian way. *(Tahbaz-Salehi and Rahnama-Rad 2010, J,Molavi, Tahbaz-Salehi 2012, Shahrampour, Rakhlin,J 2013, Rahimian, Molavi,J 2014-15), Nedic et al. (2015), Lalitha et al. 2014*

(log)-linear Learning Rules

DeGroot+Bayesian: (log)-linear aggregation

- Agents aggregate the observed beliefs of their neighbors (log)-linearly and process their private observations in a Bayesian way. *(Tahbaz-Salehi and Rahnama-Rad 2010, J,Molavi, Tahbaz-Salehi 2012, Shahrampour, Rakhlin,J 2013, Rahimian, Molavi,J 2014-15), Nedic et al. (2015), Lalitha et al. 2014*

Belief Update

$$(\log)\mu_{it+1} = a_{ii}(\log)\mathrm{BU}(\mu_{it}; \omega_{it+1}) + \sum_{j \in \mathcal{N}_i} a_{ij}(\log)\mu_{jt}$$

- (log) update can be viewed as stochastic gradient descent, applied to centralized learning
- Is there a behavioral foundation? How are these different from Bayesian? Can they be generalized?

An Axiomatic Foundation

How do individuals Aggregate beliefs?

Are these updates adhoc?

$$\mu_{it+1} = \mathrm{BU}\Big(f_{it}(\mu_i^t); \omega_{it+1}\Big),$$

- $\mu_i^t = (\mu_{j\tau})_{j \in \mathcal{N}_i, 0 \leq \tau \leq t}$: the history of beliefs of i and her neighbors
- BU : the Bayesian update, conditional on observing ω.
- $f_{it} : \Delta\Theta^{(t+1)\times|N_i|} \to \Delta\Theta$, is the *social learning rule* for agent i.
- What properties should f_{it} satisfy?
- How do these assumptions depart from Bayesian rationality?
- Under what assumptions on f_{it} is learning achieved?

Sufficient Axioms

- label neutrality: Any permutation of states $\sigma : \Theta \to \Theta$ commutes with f_{it}:
  $\text{perm}_\sigma\Big(f_{it}(\mu_i^t(\theta))\Big) = f_{it}\Big(\text{perm}_\sigma(\mu_i^t(\theta))\Big)$

- Independence of Irrelevance Alternatives (IIA): f_{it} commutes with the conditioning operator: For $\bar{\Theta} \subseteq \Theta$, $\text{cond}_{\bar{\Theta}}\left(f_{it}(\mu_i^t)\right) = f_{it}\left(\text{cond}_{\bar{\Theta}}(\mu_i^t)\right)$,
  for all histories μ_i^t: $\text{cond}_{\bar{\Theta}}(\mu_i^t) = \left(\text{cond}_{\bar{\Theta}}(\mu_{j\tau})\right)_{j \in N_i, \tau \leq t}$.
- Monotonicity: f_i is strictly monotone in $\mu_j(\theta)$ for all $j \in \mathcal{N}_i$, $\theta \in \Theta$.

- Imperfect Recall $f_{it}(\mu_i^t)$ does not depend on $\mu_{j\tau}$ for all j and all $\tau < t$ and is independent of time index t.

Axioms nail the update!

Theorem

(J., Molavi, Tahbaz-Salehi, 2015) Above axioms uniquely specify f up to a set of nonnegative constants $a_{ij} > 0$

$$\log \frac{f_{it}(\mu_i^t)(\theta)}{f_{it}(\mu_i^t)(\hat{\theta})} = \sum_{j \in N_i} a_{ij} \log \frac{\mu_{jt}(\theta)}{\mu_{jt}(\hat{\theta})}.$$

Note: LN+IIA $\rightarrow$ log-linearity. Monotonicity $\rightarrow a_{ij} \geq 0$.

Learning Rule

Corollary

If f satisfies LN, IIA, IR, and is Monotone, then the learning rule is

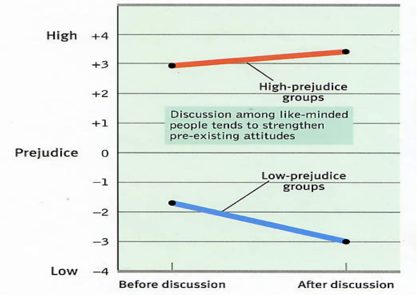
$$\log \frac{\mu_{it+1}(\theta)}{\mu_{it+1}(\hat{\theta})} = \log \frac{\ell_i^{\theta}(\omega_{it+1})}{\ell_i^{\hat{\theta}}(\omega_{it+1})} + \sum_{j \in N_i} a_{ij} \log \frac{\mu_{jt}(\theta)}{\mu_{jt}(\hat{\theta})}$$

for some $a_{ij} > 0$ when $j \in N_i$.

Log-linear update is a consequence of the behavioral axioms.

Group Polarization

- Tendency to make decisions that are more extreme than initial inclinations (Stoner'61, Lamm'75, Isenberg'86)
- Prevailing explanation: Social Comparison Theory vs. Information Influence (Persuasive Arguments theory)
- Strictly group polarizing $\rightarrow \rho(A) > 1$: Group depolarizing $\rightarrow \rho(A) < 1$. non-polarizing: $\rho(A) = 1$



Asymptotic Learning

Theorem

(J, Molavi, Tahbaz-Salehi 15) Suppose agent's social learning rules satisfy LN, M, IIA and IR.

a. *If learning rules satisfy strict group polarization, agents mislearn the state with positive probability.*
b. *If learning rules satisfy strict group depolarization, agents remain uncertain forever.*
c. *If learning rules are* non-polarizing, *all agents learn the underlying state almost surely, i.e.,* $\mu_{it}(\cdot) \longrightarrow \mathbf{1}_{\theta}(\cdot)$

- When $\rho(A) = 1$ The learning process asymptotically coincides with Bayesian learning

DeGroot Learning

- DeGroot has more deviations from Bayesian
  - Need another axiom:
  - *Separability*: Update of belief on a state is independent of beliefs on other states

Theorem

(J, Molavi, Tahbaz-Salehi 15) Suppose f_{it} satisfies LN, Monotonicity, Separability, and IR. Then the learning rule is affine: There exists constants $a_{ij} > 0$ and a_{i0} such that

$$f_{it}(\mu_i^t)(\theta) = a_{i0} + \sum_{j \in N_i} a_{ij} \mu_{jt}(\theta)$$

- Affine rule is only rule satisfying all axioms.
- Under IIA, zero belief on a state spreads, whereas here positive belief spreads

Learning under DeGroot Model

Definition (unanimity)

Social learning rule is unanimous, if

$$f_{it}(\mu, \cdots, \mu) = \mu$$

- Unanimity forces adoption of a common belief, when neighbors agree; induces a social conformation bias
- Implies a_{ij} add up to 1
- DeGroot Learning rule is non-polarizing and unanimous
- Unanimity is violated by Bayesian

Theorem

If agents' social learning rule satisfy LN, M, Separability, and IR, then all agents learn if and only if the rule is unanimous

A General Class of Learning Rules

Definition (Weak Separability)

There exists a smooth function $\psi_i : [0,1]^n \to \mathbb{R}_+$ with Elasticity of Substitution $\sigma_i^{(rj)}(x) = \frac{\psi_i^{(r)}(x)\psi_i^{(j)}(x)}{\psi_i(x)\psi_i^{(rj)}(x)} \geq 1$, ($\psi_i^{(r)} := \frac{\partial \psi_i(x_1, \cdots, x_n)}{\partial x_r}$) such that

$$\frac{f_{it}(\mu)(\theta)}{f_{it}(\mu)(\hat{\theta})} = \frac{\psi_i(\mu(\theta))}{\psi_i(\mu(\hat{\theta}))}.$$

$$\Downarrow$$

$$f_i(\mu)(\theta) = \frac{\psi_i(\mu(\theta))}{\sum_{\hat{\theta} \in \Theta} \psi_i(\mu(\hat{\theta}))}$$

- $EoS \geq 1 \Rightarrow$ Elasticity (Sensitivity) of $\{\frac{\psi_i^{(r)}(x)}{\psi_i^{(j)}(x)}$ w.r.t $\frac{x_r}{x_j}\} \leq 1 \Rightarrow$ beliefs of neighbors are *gross substitutes*.
- Generalizes IIA and Separability

Learning Under General Nonlinear Rules

$$\mu_{it+1} = \text{BU}\Big(f_{it}(\mu_i^t); \omega_{it+1}\Big) = \frac{\ell_i^\theta(\omega_{it+1}) f_i(\mu_t(\theta))}{\sum_{\hat{\theta}} \ell_i^{\hat{\theta}}(\omega_{it+1}) f_i(\mu_t(\hat{\theta}))}$$

$$= \frac{\ell_i^\theta(\omega_{it+1}) \psi_i(\mu_t(\theta))}{\sum_{\hat{\theta}} \ell_i^{\hat{\theta}}(\omega_{it+1}) \psi_i(\mu_t(\hat{\theta}))} = C(\omega_{t+1}, \mu_t(\theta))\mu_t(\theta)$$

Theorem

If agents' social learning rule satisfy Label Neutrality, Weak Separability, and Imperfect Recall, then all agents learn if ψ_i is homogeneous of degree 1, i.e., $\psi(\lambda x) = \lambda \psi_i(x)$.

special case: $\psi_i(x) = \prod_{j \in N_i} x_j^{a_{ij}}$ and $\psi_i(x) = \sum_{j \in N_i} b_{ij} x_j$, corresponding to log-linear and linear f_{it} are the two extremes

Proof Sketch

- $\mu_{i,t+1}(\theta) = C(\omega_{t+1}, \mu_t(\theta))\mu_t(\theta)$, where $[C(\omega, x)]_{ij} = \frac{\ell_i^{\theta}(\omega_i)\psi_i^{(j)}(x)}{\sum_{\hat{\theta}\in\Theta} \ell_i^{\hat{\theta}}(\omega_i)\psi_i(x)}$ is positive with ψ_i h.o.d.1
- $\psi_i(x)$ is concave and increasing
- $q_i(x) = \log \psi_i \exp(x)$ is convex and non-decreasing.
  - Equivalent to $EoS \geq 1$ condition.
- Vector function $Q(x) = \lim_{t\to\infty} \underbrace{q \circ q \circ \cdots \circ q}_{t \text{ times}}(x)$ is well-defined, continuous, non-decreasing (in each entry) and convex
- Q is additive-homogeneous: $Q(\alpha\mathbf{1} + x) = \alpha\mathbf{1} + Q(x)$
- $Q(\log \mu_t(\theta))$ is a vector submartingale that is upper bounded, and therefore converges
- $\mu_{it}(\theta) \to \mu_i^*(\theta)$ as $t \to \infty$ with $\mathbb{P}^{\theta}$-probability one, where $\mu_i^*(\theta) > 0$

General Learning Rules

- Significant generalization of previous results
- What matters for is a weak form of independence:
  - belief ratios on two states should be ratios of functions of only those two states
- functional form doesn't matter, as long as f_{it} is homogeneous and beliefs are gross substitutes
- EoS condition is equivalent to f_i or ψ_i multiplicatively convex (aka GG-convex) Niculescu 2000
- Learning rule can be any "belief-dependent" p norm-like function with $0 < p \leq 1$
- Essentially proved a random, nonlinear Perron Frobenius theorem.

Rate of Learning

Definition (total uncertainty)

Total belief on the wrong state

$$e_{it}^{\theta} = \sum_{\hat{\theta} \neq \theta} \mu_{it}(\hat{\theta})$$

Definition (Learning Rate)

$$\lambda_i^{\theta} = \liminf_{t \to \infty} \frac{1}{t} \log e_{it}$$

Rate of Learning

Definition (total uncertainty)

Total belief on the wrong state

$$e_{it}^{\theta} = \sum_{\hat{\theta} \neq \theta} \mu_{it}(\hat{\theta})$$

Definition (Learning Rate)

$$\lambda_i^{\theta} = \liminf_{t \to \infty} \frac{1}{t} \log e_{it}$$

- λ depends on
  - agents' information endowments: relative entropy
  - agents' network position: (in) and (out) eigenvector centrality.
  - Non-asymptotic rate can be derived using concentration inequalities

Relative Entropy

Definition (relative entropy)

Given $\hat{\theta} \neq \theta$,

$$h_i(\theta, \hat{\theta}) = \mathbb{E}^{\theta}\left[\log \frac{\ell_i^{\theta}(\omega)}{\ell_i^{\hat{\theta}}(\omega)}\right] = \sum_{s \in S} \ell_i^{\theta}(s) \log \frac{\ell_i^{\theta}(s)}{\ell_i^{\hat{\theta}}(s)}$$

- $h_i(\theta, \hat{\theta})$: information in favor of θ against $\hat{\theta}$ when θ is realized
- $h_i(\theta, \hat{\theta}) = 0 \quad \Rightarrow \quad$ agent i cannot distinguish θ and $\hat{\theta}$
- larger $h_i(\theta, \hat{\theta}) \quad \Rightarrow \quad$ easier to rule out $\hat{\theta}$ when θ is realized

Uniform Informativeness Order

Definition (uniform informativeness)

$$\ell_i \succeq_{\text{UI}} \ell_i'$$

if

$$h_i(\theta, \hat{\theta}) \geq h_i'(\theta, \hat{\theta}) \text{ for all } \theta, \hat{\theta}$$

- ℓ_i is more informative than ℓ_i' regardless of the realized state.
- a partial order on the set of signal structures
- weaker (more complete) than Blackwell's informativeness

Eigenvector Centrality

Definition (eigenvector centrality)

Given A, the out and in eigenvector centrality of agent i are

$$v_i = \sum_{j=1}^{n} v_j a_{ji} \quad w_i = \sum_{j=1}^{n} a_{ij} w_j$$

- a measure of the effective attention agents receive
- v_i is high when i is connected to central agents.
- well-defined and positive given a strongly connected network
- in the general case, v is the Perron-Frobenius *nonlinear eigenvector* of the monotone, homogeneous map $f = [f_1, \cdots, f_n]'$

Rate of Learning

Theorem

Suppose the social learning rule satisfies LN,IIA, Monotonicity, and is non-polarizing

(a) $\lambda_i \in (0, +\infty) \quad \Rightarrow \quad$ *exponential convergence*

(b) $\lambda_i^\theta = w_i \min_{\hat{\theta} \neq \theta} \sum_{j=1}^n v_j h_j(\theta, \hat{\theta})$

- We use λ_i^θ as the rate of learning of agent i.
- Zeroth order term of rate depends on spectral gap

Remarks

When learning rule is unanimous

$$\lambda_i^\theta = \min_{\hat{\theta}\neq\theta} \sum_{i=1}^{n} v_i h_i(\theta, \hat{\theta})$$

- A uniform increase in information increases the rate:

$$\ell_i \succeq_{\text{UI}} \ell'_i \qquad \forall i \quad \Rightarrow \quad r \geq r'$$

- The rate is determined by $\theta, \hat{\theta}$ that are hardest to distinguish.

- The network and information effects are not decoupled.

  $\Rightarrow$ The interplay of the two has non-trivial implications.

Comparative Analysis

- The rate of learning is sensitive to the allocation of signals.
- Fix the total information but reallocate it among agents.

Comparative Analysis

- The rate of learning is sensitive to the allocation of signals.
- Fix the total information but reallocate it among agents.

Definition (Reallocation)

ℓ is a reallocation of ℓ' if there exists a permutation σ such that

$$\ell_i = \ell'_{\sigma(i)} \qquad \forall i$$

Under which allocation of signals is learning the fastest?

Learning Under Uniform Informativeness

Proposition

Suppose

- *agents' signals are comparable with respect to* $\succeq_{UI}$*;*
- $\ell_i \succeq_{UI} \ell_j$ *if and only if* $v_i \geq v_j$.

Then, no reallocation of signals increases the rate of learning.

- Positive assortative matching of centralities and signal qualities maximizes the rate of learning.
- Intuition: Irrespective of the realized state, the most informative signals receive the most attention.

Other Cases

- The information ordering which leads to optimality of positive assortative matching is strong.
- Other orderings could lead to dramatically different results.

In fact, the optimal allocation could correspond to the least central agent making the "best" observations.

Experts: Apples vs. Oranges An Example (1/2)

- $\Theta = \{\theta_0, \theta_1, \ldots, \theta_n\}$
- $S = \{\text{Head}, \text{Tail}\}$
- $\pi_i > \frac{1}{2}$
- Agent i is the expert in θ_i.
- Nobody else can distinguish θ_i from θ_0.

$$\ell_i^\theta(s) : \begin{array}{c} \\ \theta_0 \\ \theta_1 \\ \\ \theta_i \\ \\ \theta_n \end{array} \begin{array}{c} \begin{array}{cc} \text{Head} & \text{Tail} \end{array} \\ \begin{pmatrix} 1-\pi_i & \pi_i \\ 1-\pi_i & \pi_i \\ \vdots & \vdots \\ \pi_i & 1-\pi_i \\ \vdots & \vdots \\ 1-\pi_i & \pi_i \end{pmatrix} \end{array}$$

Experts: Apples vs. Oranges An Example (1/2)

- $\Theta = \{\theta_0, \theta_1, \ldots, \theta_n\}$
- $S = \{\text{Head}, \text{Tail}\}$
- $\pi_i > \frac{1}{2}$
- Agent i is the expert in θ_i.
- Nobody else can distinguish θ_i from θ_0.

$$\ell_i^{\theta}(s) : \quad \begin{array}{c} \\ \theta_0 \\ \theta_1 \\ \vdots \\ \theta_i \\ \vdots \\ \theta_n \end{array} \begin{array}{c} \begin{array}{cc} \text{Head} & \text{Tail} \end{array} \\ \begin{pmatrix} 1-\pi_i & \pi_i \\ 1-\pi_i & \pi_i \\ \vdots & \vdots \\ \pi_i & 1-\pi_i \\ \vdots & \vdots \\ 1-\pi_i & \pi_i \end{pmatrix} \end{array}$$

$$r = \min_i v_i \underbrace{h_i(\theta_i, \theta_0)}_{H_i},$$

- The rate of learning is determined by the agent with the smallest centrality × expertise.

Experts: An Example (2/2)

Proposition

Suppose that $H_i \geq H_j$ if and only if $v_i \leq v_j$.
Then, no reallocation of signals increases the rate of learning.

Experts: An Example (2/2)

Proposition

Suppose that $H_i \geq H_j$ if and only if $v_i \leq v_j$.
Then, no reallocation of signals increases the rate of learning.

- Negative assortative matching of centralities and signal qualities maximizes the rate of learning.
- In the optimal allocation, the least central agent has the highest expertise.
- There needs to be a mismatch between two bottlenecks:
  - network bottleneck effect: Peripheral agents receive low attention.
  - identification bottleneck effect: Some states are harder to tell apart.

Experts (1/2)

- relative informativeness of agent i's signals for $(\theta, \hat{\theta})$:

$$\gamma_i(\theta, \hat{\theta}) = \sup\{\beta : h_i(\theta, \hat{\theta}) \geq \beta h_j(\theta, \hat{\theta}) \quad \forall j \neq i\}$$

- specialty of agent i:

$$E_i = \{(\theta, \hat{\theta}) : \theta \neq \hat{\theta} \ \text{ and } \ \gamma_i(\theta, \hat{\theta}) \geq 1\}$$

Definition (expertise)

- relative expertise: $\gamma_i = \min\{\gamma_i(\theta, \hat{\theta}) : (\theta, \hat{\theta}) \in E_i\}$
- absolute expertise: $\varepsilon_i = \min\{h_i(\theta, \hat{\theta}) : (\theta, \hat{\theta}) \in E_i\}$

Experts (2/2)

Proposition

Suppose that

- $E_i \neq \emptyset$ *for all* i;
- $\varepsilon_i \geq \varepsilon_j$ *if and only if* $v_i \leq v_j$.

Then, reallocations of signals do not increase the rate by more than $\alpha(\max_i \varepsilon_i)/(\min_i \gamma_i)$.

- 1st condition: Agents are all experts.
- 2nd condition: The least central agents have the highest absolute expertise.

Network Regularity

Definition (regularity)

$$A \quad \succeq_{\text{reg}} \quad A'$$

if

$$\sum_{i=1}^{k} v_{[i]} \quad \leq \quad \sum_{i=1}^{k} v'_{[i]} \quad \forall k = 1, \cdots, n-1$$

$$\Updownarrow$$

$$v^{\downarrow} \quad \text{FOSD}^{a} \quad v'^{\downarrow}$$

[a] first-order stochastically dominates

Network Regularity and Learning

- Consider the rate under the best allocation: r^*.

Is r^* higher for regular or irregular networks?

Network Regularity and Learning

- Consider the rate under the best allocation: r^*.

Is r^* higher for regular or irregular networks?

Proposition

Suppose agents' signals are comparable with respect to $\succeq_{UI}$. Then,

$$A \quad \succeq_{reg} \quad A' \qquad \Rightarrow \qquad r^* \leq r'^*$$

Network Regularity and Learning

- Consider the rate under the best allocation: r^*.

Is r^* higher for regular or irregular networks?

Proposition

Suppose agents' signals are comparable with respect to $\succeq_{UI}$. Then,

$$A \quad \succeq_{reg} \quad A' \qquad \Rightarrow \qquad r^* \leq r'^*$$

- The gap can grow unboundedly in large networks

Network Regularity and Learning: An Example

- $\Theta = \{\theta_0, \theta_1, \ldots, \theta_n\}$
- $S = \{\text{Head}, \text{Tail}\}$
- $\pi > \frac{1}{2}$

$$\ell_i^\theta(s) : \begin{array}{c} \\ \theta_0 \\ \theta_1 \\ \\ \theta_i \\ \\ \theta_n \end{array} \begin{array}{cc} \text{Head} & \text{Tail} \\ \left(\begin{array}{cc} 1-\pi & \pi \\ 1-\pi & \pi \\ \vdots & \vdots \\ \pi & 1-\pi \\ \vdots & \vdots \\ 1-\pi & \pi \end{array}\right) \end{array}$$

Proposition

$$A \quad \succeq_{reg} \quad A' \qquad \Rightarrow \qquad r^* \geq r'^*$$

- Ordering of networks is reversed with expert agents!
- The gap does not grow unboundedly.

$\Rightarrow$ Rates of learning in all large networks are similar.

Conclusion

- Developed an axiomatic foundation for non-Bayesian learning
- Key deviations: imperfect recall, IIA, and monotonicity
- Rather than postulate a functional form, derive them from axioms
- The interplay of information and network structures determines information aggregation.
- When the information content fully ordered , optimal allocation of signals entails central agents to receive better signals.
- Using different notions of information ordering can result in significantly different conclusions.
- In the optimal allocation, the least central agents could make the best observations.
- Model generalizes to very general nonlinear updates

Summary of the Results

- Developed a foundation for a general family of non-Bayesian models with
  - continuous flow of new information
  - heterogenous observations
  - asymptotic agreement with the Bayesian benchmark
  - clear behavioral implications as compared to Bayesian learning

Summary of the Results

- Developed a foundation for a general family of non-Bayesian models with
  - continuous flow of new information
  - heterogenous observations
  - asymptotic agreement with the Bayesian benchmark
  - clear behavioral implications as compared to Bayesian learning

- A characterization of the rate learning in terms of
  - social network structure: eigenvector centrality
  - agents' information endowments: relative entropy

Summary of the Results

- Developed a foundation for a general family of non-Bayesian models with
  - continuous flow of new information
  - heterogenous observations
  - asymptotic agreement with the Bayesian benchmark
  - clear behavioral implications as compared to Bayesian learning
- A characterization of the rate learning in terms of
  - social network structure: eigenvector centrality
  - agents' information endowments: relative entropy
- Under which information allocation learning is the fastest?
  - comparable info endowments $\Rightarrow$ positive assortative matching
  - complimentary info endowments $\Rightarrow$ negative assortative matching

Summary of the Results

- Developed a foundation for a general family of non-Bayesian models with
  - continuous flow of new information
  - heterogenous observations
  - asymptotic agreement with the Bayesian benchmark
  - clear behavioral implications as compared to Bayesian learning
- A characterization of the rate learning in terms of
  - social network structure: eigenvector centrality
  - agents' information endowments: relative entropy
- Under which information allocation learning is the fastest?
  - comparable info endowments $\Rightarrow$ positive assortative matching
  - complimentary info endowments $\Rightarrow$ negative assortative matching
- Comparative analysis of networks in terms of regularity